arXiv:1501.06641v1 [math.PR] 27 Jan 2015

On singular values distribution of a large auto-covariance matrix in the ultra-dimensional regime

Qinwen Wang and Jianfeng Yao*

Qinwen Wang
Department of Mathematics
Zhejiang University
e-mail: wqw8813@gmail.com

Jianfeng Yao
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam, Hong Kong
e-mail: jeffyao@hku.hk

Abstract: Let $(\varepsilon_t)_{t>0}$ be a sequence of independent real random vectors of p-dimension and let $X_T = \sum_{t=s+1}^{s+T} \varepsilon_t \varepsilon_{t-s}^T / T$ be the lag-s (s is a fixed positive integer) auto-covariance matrix of ε_t. This paper investigates the limiting behavior of the singular values of X_T under the so-called *ultra-dimensional regime* where $p \to \infty$ and $T \to \infty$ in a related way such that $p/T \to 0$. First, we show that the singular value distribution of X_T after a suitable normalization converges to a nonrandom limit G (quarter law) under the forth-moment condition. Second, we establish the convergence of its largest singular value to the right edge of G. Both results are derived using the moment method.

AMS 2000 subject classifications: 15A52, 60F15;.
Keywords and phrases: Auto-covariance matrix, Singular values, Limiting spectral distribution, Ultra-dimensional data, Largest eigenvalue, Moment method.

*The research of J. Yao is partly supported by GRF Grant HKU 705413P.

1. Introduction

Let s be a fixed positive integer and $(\varepsilon_t)_{1\le t\le T+s}$ a sequence of independent real random vectors, where $\varepsilon_t = (\varepsilon_{it})_{1\le i\le p}$ has independent coordinates satisfying $\mathbb{E}\varepsilon_{it} = 0$ and $\mathbb{E}\varepsilon_{it}^2 = 1$. Consider the so-called *lag-s sample autocovariance matrix* of (ε_t) defined as

$$X_T = \frac{1}{T}\sum_{t=s+1}^{s+T} \varepsilon_t \varepsilon_{t-s}^T \ . \tag{1.1}$$

Motivated by their application in high-dimensional statistical analysis where the dimensions p and T are assumed large (tending to infinity), spectral analysis of such sample autocovariance matrices have attracted much attention in recent literature in random matrix theory. For example, perturbation theory on the matrix X_T has been carried out in Lam and Yao (2012) and Li et al. (2014) for estimating the number of factors in a large dimensional factor model of type

$$y_t = \Lambda f_t + \varepsilon_t + \mu \ , \tag{1.2}$$

where $\{y_t\}$ is a p-dimensional sequence observed at time t, $\{f_t\}$ a sequence of m-dimensional "latent factor" ($m \ll p$) uncorrelated with the error process $\{\varepsilon_t\}$ and $\mu \in \mathbb{R}^p$ is the general mean. Since X_T is not symmetric, its spectral distribution is given by the set of its singular values which are by definition the square roots of positive eigenvalues of

$$A_T := X_T X_T^T \ . \tag{1.3}$$

To our best knowledge, all the existing results on X_T (or A_T) are found under what we will refer as the *Marčenko-Pastur regime*, or simply the *MP regime*, where

$$p \to \infty, \quad T \to \infty \quad \text{and} \quad p/T \to c > 0 \ . \tag{1.4}$$

For example, Jin et al (2014) derives the limit of the eigenvalue distributions (ESD) of the symmetrized auto-covariance matrix $\frac{1}{2}(X_T + X_T^T)$; and Wang et al. (2013) establishes the exact separation property of the ESD which also implies the convergence of its extreme

eigenvalues. For the singular value distribution of X_T, the limit (LSD) has been established in Li et al. (2013) using the method of Stieltjes transform and in Wang and Yao (2014) using the moment method. The latter paper also establishes the almost sure convergence of the largest singular value of X_T to the right edge of the LSD, thanks to the moment method. Related results are also proposed in Liu et al. (2013) where the sequence (ε_t) is replaced by a more general time series.

In this paper, we investigate the same questions as in Wang and Yao (2014) but under a different asymptotic regime, the so-called *ultra-dimensional regime* where

$$p \to \infty, \quad T \to \infty \quad \text{and} \quad p/T \to 0. \tag{1.5}$$

It is naturally expected that the limit under this regime will be much different than under the MP regime above. The findings of the paper confirm this difference by providing a new limit of the singular value distribution of X_T under the ultra-dimensional regime.

In a related paper Wang et Paul (2014), the authors also adopted the ultra-dimensional regime to derive the LSD for a large class of separable sample covariance matrices. However, the autocovariance matrix X_T considered in this paper is very different of these separable sample covariance matrices.

Recalling the definition of A_T in (1.3), we have

$$A_T(i,j) = \frac{1}{T^2} \sum_{l=1}^{p} \sum_{m=1}^{T} \sum_{n=1}^{T} \varepsilon_{i\,m+s} \varepsilon_{lm} \varepsilon_{j\,n+s} \varepsilon_{ln} \ .$$

It follows by simple calculations that

$$\mathbb{E} A_T(i,j) = \begin{cases} 0, & i \neq j \ , \\ p/T, & i = j \ , \end{cases}$$

and for $i \neq j$,

$$\operatorname{Var} A_T(i,j) = \mathbb{E} A_T^2(i,j) = \frac{p}{T^2} \ .$$

The row sum of the variances $\mathrm{Var}\, A_T(i,j)$ is thus of order p^2/T^2. Therefore, in order to have the spectrum of A_T be of constant order when $p/T \to 0$, we should normalise it as

$$A := \frac{A_T}{\sqrt{p^2/T^2}} = \frac{T}{p} X_T X_T^T \ . \tag{1.6}$$

The main results of the paper are as follows. First in Section 2, we derive the almost sure limit of the singular value distribution of $\sqrt{\frac{T}{p}} X_T$ under the ultra-dimensional regime and assuming that the fourth moment of the entries $\{\varepsilon_{it}\}$ are uniformly bounded. This limit (LSD) simply equals to the image measure of the semi-circle law on $[-2,2]$ by the absolute value transformation $x \mapsto |x|$. Next in Section 3, we establish the almost sure convergence of the largest singular value of $\sqrt{\frac{T}{p}} X_T$ to 2 assuming that the entries $\{\varepsilon_{it}\}$ has a uniformly bounded moment of order $4+\nu$ for some $\nu > 0$. Both results are derived using the moment method. Some technical details on the traditional truncation and renormalisation steps are postponed to the appendixes.

2. Limiting spectral distribution by the moment method

In this section, we show that when $p/T \to 0$, the ESD of the singular values of $\sqrt{\frac{T}{p}} X_T$ tends to a nonrandom limit, which is linked to the well known semi-circle law.

Theorem 2.1. Suppose the following conditions hold:

(a). $(\varepsilon_t)_t$ is a sequence of independent p-dimensional real valued random vectors with independent entries ε_{it}, $1 \le i \le p$, satisfying

$$\mathbb{E}(\varepsilon_{it}) = 0, \quad \mathbb{E}\varepsilon_{it}^2 = 1, \quad \sup_{it} \mathbb{E}(\varepsilon_{it}^4) < \infty \ . \tag{2.1}$$

(b). Both p and T tend to infinity in a related way such that $p/T \to 0$.

Then, with probability one, the empirical distribution of the singular values of $\sqrt{\frac{T}{p}} X_T$ tends to the quarter law G with density function

$$g(x) = \frac{1}{\pi}\sqrt{4 - x^2} \ , \qquad 0 < x \le 2 \ . \tag{2.2}$$

Remark 2.1. Recall that the quarter law G is the image measure of the semi-circle law by the absolute value transformation. It is also worth noticing that if there were no lag, i.e. $s = 0$, the matrix X_T would be a standard sample covariance matrix; and in this case the spectral distribution of $\sqrt{\frac{T}{p}}(X_T - I_p)$ would converge to the semi-circle law, see Bai and Yin (1988). The case of a auto-covariance matrix X_T with a positive lag $s > 0$ is then very different.

Since the singular values of $\sqrt{\frac{T}{p}}X_T$ are the square roots of the eigenvalues of $\frac{T}{p}X_T X_T^T$, in the remaining of this paper, we focus on the limiting behaviours of the eigenvalues of $\frac{T}{p}X_T X_T^T$. These properties can then be transferred to the singular values of $\sqrt{\frac{T}{p}}X_T$ by the square-root transformation $x \mapsto \sqrt{x}$.

Theorem 2.2. Under the same conditions as in Theorem 2.1, with probability one, the empirical spectral distribution F^A of the matrix A in (1.6) tends to a limiting distribution F, which is the image measure of the semi-circle law on $[-2, 2]$ by the square transformation. In particular, its k-th moment is:

$$m_k = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} , \tag{2.3}$$

and its Stieltjes transform $s(z)$ and density function $f(x)$ are given by

$$s(z) = -\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{z}} , \qquad z \notin (0, 4] , \tag{2.4}$$

and

$$f(x) = \frac{1}{\pi}\sqrt{\frac{1}{x} - \frac{1}{4}} , \qquad 0 < x \leq 4 , \tag{2.5}$$

respectively.

Remark 2.2. The k-th moment in (2.3) is exactly the $2k$-th moment of the LSD of a standard Wigner matrix, which is also the number of Dyck paths of length $2k$ (for the definition of Dyck paths, we refer to Tao (2012)). Notice also that the density function f is unbounded at the origin.

The remaining of the section is devoted to the proof of Theorem 2.2 using the moment method. The k-th moment of the ESD F^A of A is

$$m_k(A) = \frac{1}{p} \operatorname{tr} A^k = \sum_{\mathbf{i}=1}^{T} \sum_{\mathbf{j}=1}^{p} \frac{1}{p^{k+1}T^k} \varepsilon_{j_1\, i_1} \varepsilon_{j_1\, i_2} \varepsilon_{j_2\, s+i_2} \varepsilon_{j_2\, s+i_3} \cdots$$

$$\varepsilon_{j_{2k-1}\, i_{2k-1}} \varepsilon_{j_{2k-1}\, i_{2k}} \varepsilon_{j_{2k}\, s+i_{2k}} \varepsilon_{j_{2k}\, s+i_1} \ . \tag{2.6}$$

Here, the indexes in $\mathbf{i} = (i_1, \cdots, i_{2k})$ run over $1, 2, \cdots, T$ and the indexes in $\mathbf{j} = (j_1, \cdots, j_{2k})$ run over $1, 2, \cdots, p$.

The core of the proof is to establish the following two assertions:

$$\text{(I).} \quad \mathbb{E} m_k(A) \to m_k = \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} , \quad k \geq 0;$$

$$\text{(II).} \quad \sum_{p=1}^{\infty} \operatorname{Var}(m_k(A)) < \infty \ .$$

This is given in the Subsections 2.1, 2.2 and 2.3 below. It follows from these assertions that almost surely, $m_k(A) \to m_k$ for all $k \geq 0$. Since the limiting moment sequence (m_k) clearly satisfies the Carleman's condition, i.e. $\sum_{k>0} m_{2k}^{-1/(2k)} = \infty$, we deduce that almost surely, the sequence of ESDs F^A weakly converges to a probability measure F whose moments are exactly (m_k). Next, notice that m_k is exactly the number of Dyck paths of length $2k$ (Tao , 2012), which is also the $2k$-th moment of the semi-circle law with support $[-2, 2]$, it follows that the LSD F equals to the image of the semi-circle law by the square transformation $x \to x^2$. The formula in (2.4) and (2.5) are thus easily derived and the proof of Theorem 2.2 is complete.

2.1. Preliminary steps and some graph concepts

We now introduce the proofs for Assertions (I) and (II). First we show that with a uniformly bounded fourth order moment, the variables $\{\varepsilon_{it}\}$ can be truncated at rate $\eta T^{1/4}$ for some vanishing sequence $\eta = \eta(T)$. This is justified in Appendix A. After these

truncation, centralisation and rescaling steps, we may assume in all the following that

$$\mathbb{E}(\varepsilon_{ij}) = 0,\ \mathbb{E}\varepsilon_{ij}^2 = 1,\ |\varepsilon_{ij}| \le \eta T^{1/4}\ , \tag{2.7}$$

where η is chosen such that $\eta \to 0$ but $\eta T^{1/4} \to \infty$.

Now we introduce some basic concepts for graphs associated to the big sum in (2.6). Let

$$\begin{gathered}\psi(e_1, \cdots, e_m) := \text{number of distinct entities among } e_1, \cdots, e_m\ ,\\ \mathbf{i} := (i_1, \cdots, i_{2k}), \quad \mathbf{j} := (j_1, \cdots, j_{2k}),\\ 1 \le i_a \le T, \quad 1 \le j_b \le p, \quad a, b = 1, \cdots, 2k,\\ A(t,s) := \{(\mathbf{i}, \mathbf{j}) : \psi(\mathbf{i}) = t, \psi(\mathbf{j}) = s\}\ .\end{gathered}$$

Define $Q(i,j)$ as the multigraph as follows: Let I-line, J-line be two parallel lines, plot $i_1, \cdots, i_{2k}$ on the I-line, $j_1, \cdots, j_{2k}$ on the J-line, called the *I-vertexes* and *J-vertexes*, respectively. Draw k down edges from i_{2u-1} to j_{2u-1}, k down edges from $i_{2u} + s$ to j_{2u}, k up edges from j_{2u-1} to i_{2u}, k up edges from j_{2u} to $i_{2u+1} + s$ (all these up and down edges are called *vertical edges*) and k *horizontal edges* from i_{2u} to $i_{2u} + s$, k horizontal edges from $i_{2u-1} + s$ to i_{2u-1} (with the convention that $i_{2k+1} = i_1$), where all the u's are in the region: $1 \le u \le k$. An example of the multi-graph $Q(i,j)$ with $k = 3$ is presented in the following Figure 1.

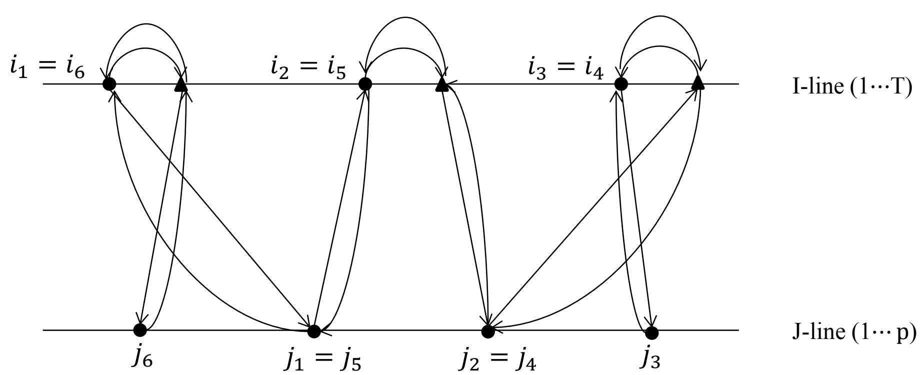

Figure 1: An example of the multigraph $Q(i,j)$ with $k = 3$.

In the graph $Q(i,j)$, once a I-vertex i_l is fixed, so is $i_l + s$. For this reason, we glue all the I-vertexes which are connected through horizon edges and denote the resulting graph

as $M(A(t,s))$, where $A(t,s)$ is the index set that has t distinct I-vertexes and s distinct J-vertexes. An example of $M(A(3,4))$ that corresponds to the $Q(i,j)$ in Figure 1 is presented in the following Figure 2.

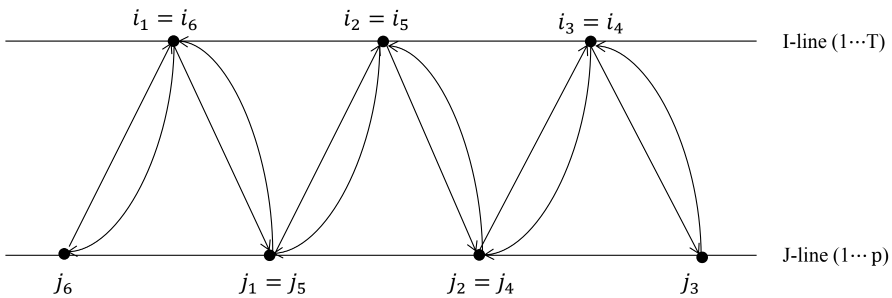

Figure 2: An example of $M(A(3,4))$ that corresponds to the $Q(i,j)$ in Figure 1.

2.2. Proof of Assertion (I)

Recall the expression of $m_k(A)$ in (2.6), we have

$$\begin{aligned}
\mathbb{E}m_k(A) &= \sum_{\mathbf{i}=1}^{T}\sum_{\mathbf{j}=1}^{p}\frac{1}{p^{k+1}T^k}\mathbb{E}\big[\varepsilon_{j_1\,i_1}\varepsilon_{j_1\,i_2}\varepsilon_{j_2\,s+i_2}\varepsilon_{j_2\,s+i_3}\varepsilon_{j_3\,i_3}\varepsilon_{j_3\,i_4}\varepsilon_{j_4\,s+i_4}\varepsilon_{j_4\,s+i_5} \\
&\qquad\cdots\varepsilon_{j_{2k-1}\,i_{2k-1}}\varepsilon_{j_{2k-1}\,i_{2k}}\varepsilon_{j_{2k}\,s+i_{2k}}\varepsilon_{j_{2k}\,s+i_1}\big] \\
&= \sum_{t,s}\frac{1}{p^{k+1}T^k}\sum_{M(A(t,s))}p(p-1)\cdots(p-s+1)T(T-1)\cdots(T-t+1) \\
&\qquad\cdot\mathbb{E}\big[\varepsilon_{j_1\,i_1}\varepsilon_{j_1\,i_2}\varepsilon_{j_2\,s+i_2}\varepsilon_{j_2\,s+i_3}\cdots\varepsilon_{j_{2k-1}\,i_{2k}}\varepsilon_{j_{2k}\,s+i_{2k}}\varepsilon_{j_{2k}\,s+i_1}\big] \\
&:= \sum_{t,s}S(t,s)\ , \qquad (2.8)
\end{aligned}$$

where

$$\begin{aligned}
S(t,s) &= \frac{1}{p^{k+1}T^k}\sum_{M(A(t,s))}p(p-1)\cdots(p-s+1)T(T-1)\cdots(T-t+1) \\
&\qquad\cdot\mathbb{E}\big[\varepsilon_{j_1\,i_1}\varepsilon_{j_1\,i_2}\cdots\varepsilon_{j_{2k}\,s+i_{2k}}\varepsilon_{j_{2k}\,s+i_1}\big]\ . \qquad (2.9)
\end{aligned}$$

Then we assert a lemma stating that $|S(t,s)| \to 0$ except for one particular term.

Lemma 2.1. $|S(t,s)| \to 0$ as $p \to \infty$ unless $t = k$ and $s = k+1$.

Suppose Lemma 2.1 holds true for a moment, then according to (2.8) and (2.9), we have

$$\mathbb{E} m_k(A) = S(k, k+1) + o(1) = \mathbb{E}[\cdot] \cdot \#\{M(A(k, k+1))\} + o(1) , \tag{2.10}$$

where $\mathbb{E}[\cdot]$ refers to the expectation part in (2.9) and $\#\{M(A(k,k+1))\}$ refers to the number of isomorphism class that have k distinct I-vertexes and $k+1$ distinct J-vertexes. First, we show the expectation part $\mathbb{E}[\cdot]$ equals 1 when $t = k$ and $s = k+1$. Let v_m denote the number of edges in $M(A(t,s))$ whose degree is m. Then we have the total number of edges having the following relationship:

$$v_1 + 2v_2 + \cdots + 4kv_{4k} = 4k . \tag{2.11}$$

Since we have $\mathbb{E}\varepsilon_{ij} = 0$ in (2.7), all the multiplicities of the edges in the graph $M(A(t,s))$ should be at least two, that is $v_1 = 0$. On the other hand, $M(A(t,s))$ is a connected graph with $t+s$ vertexes and $v_1 + \cdots + v_{4k}$ $(= v_2 + \cdots + v_{4k})$ edges, we have when $t = k$ and $s = k+1$:

$$\begin{aligned} 2k+1 = t+s &\le v_1 + \cdots + v_{4k} + 1 = v_2 + \cdots + v_{4k} + 1 \\ &\le \frac{1}{2}(2v_2 + 3v_3 + \cdots + 4kv_{4k}) + 1 = 2k+1 , \end{aligned} \tag{2.12}$$

where the last equality is due to (2.11) with $v_1 = 0$. Then we have all the inequalities in (2.12) become equalities, that is,

$$v_2 + \cdots + v_{4k} + 1 = \frac{1}{2}(2v_2 + 3v_3 + \cdots + 4kv_{4k}) + 1 = 2k+1 ,$$

which leads to the fact that

$$v_3 = v_4 = \cdots = v_{4k} = 0, \quad v_2 = 2k . \tag{2.13}$$

This means that all the edges in the graph $M(A(k,k+1))$ is repeated exactly twice, so the part of expectation

$$\mathbb{E}\big[\varepsilon_{j_1\, i_1}\varepsilon_{j_1\, i_2} \cdots \varepsilon_{j_{2k}\, s+i_{2k}}\varepsilon_{j_{2k}\, s+i_1}\big] = \big(\mathbb{E}\varepsilon_{ji}^2\big)^{2k} = 1 . \tag{2.14}$$

Second, the number of isomorphism class in $M(A(t,s))$ (with each edge repeated at least twice in the original graph $Q(i,j)$) is given by the notation $f_{t-1}(k)$ in Wang and Yao (2014), where

$$f_{t-1}(k) = \frac{1}{k}\begin{pmatrix} 2k \\ t-1 \end{pmatrix}\begin{pmatrix} k \\ t \end{pmatrix} .$$

Therefore, in this special case when $t = k$ and $s = k+1$, we have

$$\#\{M(A(k,k+1))\} = f_{k-1}(k) = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} . \tag{2.15}$$

Finally, combine (2.10), (2.14) and (2.15), we have

$$\mathbb{E}m_k(A) = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} + o(1) .$$

Assertion (I) is then proved.

It remains to prove Lemma 2.1.

Proof. (of Lemma 2.1) Denote b_l as the degree that associated to the I-vertex i_l $(1 \le l \le t)$ in $M(A(t,s))$, then we have $b_1 + \cdots + b_t = 4k$, which is the total number of edges. On the other hand, since each edge in $M(A(t,s))$ is repeated at least twice (otherwise, there exist at least one single edge, so the expectation will be zero), we have each degree b_l at least four (we glue the original I-vertexes i_l and $i_l + s$ in $M(A(t,s))$). Therefore, we have

$$4k = b_1 + \cdots + b_t \ge 4t ,$$

which is $t \le k$.

Now, consider the following two cases separately.

Case 1: $s > k+1$**.**

Recall the definition of v_m in (2.11), which satisfies that

$$v_1 + 2v_2 + \cdots + 4kv_{4k} = 2v_2 + \cdots + 4kv_{4k} = 4k$$

and

$$t + s \le v_1 + \cdots + v_{4k} + 1 = v_2 + \cdots + v_{4k} + 1 \ .$$

We can bound the expectation part as follows:

$$\begin{aligned}
&\left|\mathbb{E}[\varepsilon_{j_1\, i_1}\varepsilon_{j_1\, i_2}\varepsilon_{j_2\, s+i_2}\varepsilon_{j_2\, s+i_3}\cdots\varepsilon_{j_{2k}\, s+i_{2k}}\varepsilon_{j_{2k}\, s+i_1}]\right| \\
&\le \left|\mathbb{E}\varepsilon_{ji}^2\right|^{v_2}\cdots\left|\mathbb{E}\varepsilon_{ji}^{4k}\right|^{v_{4k}} \le \left(\eta T^{1/4}\right)^{v_3+2v_4+\cdots+(4k-2)v_{4k}} \\
&= \left(\eta T^{1/4}\right)^{3v_3+4v_4+\cdots+4kv_{4k}-2(v_3+v_4+\cdots+v_{4k})} \\
&= \left(\eta T^{1/4}\right)^{4k-2(v_2+v_3+\cdots+v_{4k})} \le \left(\eta T^{1/4}\right)^{4k-2(t+s-1)} \ . \qquad (2.16)
\end{aligned}$$

Then we have according to (2.9) that

$$\begin{aligned}
|S(t,s)| &\le \frac{1}{p^{k+1}T^k}T^t p^s \left(\eta T^{1/4}\right)^{4k-2(t+s-1)} \#\{M(A(t,s))\} \\
&= \frac{p^{s-k-1}}{T^{\frac{1}{2}(s-t-1)}}\eta^{4k-2(t+s-1)}\#\{M(A(t,s))\} \\
&= O\left(\frac{p^{s-k-1}}{T^{\frac{1}{2}(s-t-1)}}\eta^{4k-2(t+s-1)}\right) \ , \qquad (2.17)
\end{aligned}$$

where the last equality is due to the fact that $\#\{M(A(t,s))\}$ is a function of k (k is fixed), which could be bounded by a large enough constant.

Since $s > k+1$ and $t+s-1 \le 2k$, then

$$s-k-1-\frac{s}{2}+\frac{t}{2}+\frac{1}{2} = \frac{s}{2}-k+\frac{t}{2}-\frac{1}{2} = \frac{1}{2}(s+t-2k-1) \le 0 \ ,$$

which is

$$0 < s-k-1 \le \frac{1}{2}(s-t-1) \ .$$

So, (2.17) reduces to

$$|S(t,k)| \le O\left(\left(\frac{p}{T}\right)^{\frac{s-k-1}{2}}\eta^{4k-2(t+s-1)}\right) \to 0 \ , \qquad (2.18)$$

which is due to the fact that $s-k-1>0$ and $p/T \to 0$.

Case 2: $s \leq k+1$, but not $t=k$ and $s=k+1$.

For the same reason as before, we have t distinct I-vertexes, each degree is at least four, so we have another estimation for the expectation part:

$$\left|\mathbb{E}[\varepsilon_{j_1\, i_1}\varepsilon_{j_1\, i_2}\varepsilon_{j_2\, s+i_2}\varepsilon_{j_2\, s+i_3}\cdots\varepsilon_{j_{2k}\, s+i_{2k}}\varepsilon_{j_{2k}\, s+i_1}]\right| \leq \left(\eta T^{1/4}\right)^{4k-4t} \ . \tag{2.19}$$

Therefore,

$$\begin{aligned} |S(t,s)| &\leq \frac{1}{p^{k+1}T^k} T^t p^s \left(\eta T^{1/4}\right)^{4k-4t} \#\{M(A(t,s))\} \\ &= O\left(\frac{\eta^{4k-4t}}{p^{k+1-s}}\right) \ , \end{aligned} \tag{2.20}$$

which is also due to the fact that $\#\{M(A(t,s))\} = O(1)$.

Case 2 contains three situations:

$$\begin{aligned} &(1).\ t=k \text{ and } s<k+1: |S(t,s)| \leq O\left(\frac{1}{p^{k+1-s}}\right) \to 0 \ ; \\ &(2).\ t<k \text{ and } s=k+1: |S(t,s)| \leq O\left(\eta^{4k-4t}\right) \to 0 \ ; \\ &(3).\ t<k \text{ and } s<k+1: |S(t,s)| \leq O\left(\frac{\eta^{4k-4t}}{p^{k+1-s}}\right) \to 0 \ . \end{aligned} \tag{2.21}$$

Combine (2.18) and (2.21), we have $|S(t,k)| \to 0$ as $p \to \infty$ unless

$$\begin{cases} t = k \\ s = k+1 \ . \end{cases}$$

□

2.3. Proof of Assertion (II)

Recall

$$\begin{aligned} &\mathrm{Var}(m_k(A)) \\ &= \frac{1}{p^{2k+2}T^{2k}} \sum_{\mathbf{i_1},\mathbf{j_1},\mathbf{i_2},\mathbf{j_2}} \left[\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right) - \mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\right)\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)\right] \ . \end{aligned} \tag{2.22}$$

If $Q(\mathbf{i_1},\mathbf{j_1})$ has no edges coincident with edges of $Q(\mathbf{i_2},\mathbf{j_2})$, then

$$\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)-\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\right)\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)=0$$

by independence between $\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}$ and $\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}$. If $Q=Q(\mathbf{i_1},\mathbf{j_1})\bigcup Q(\mathbf{i_2},\mathbf{j_2})$ has an overall single edge, then

$$\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)=\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\right)\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)=0\ ,$$

so in the above two cases, we have $\mathrm{Var}(m_k(A))=0$.

Now, suppose $Q=Q(\mathbf{i_1},\mathbf{j_1})\bigcup Q(\mathbf{i_2},\mathbf{j_2})$ has no single edge, $Q(\mathbf{i_1},\mathbf{j_1})$ and $Q(\mathbf{i_2},\mathbf{j_2})$ have common edges. Let the number of vertexes of $Q(\mathbf{i_1},\mathbf{j_1})$, $Q(\mathbf{i_2},\mathbf{j_2})$, $Q=Q(\mathbf{i_1},\mathbf{j_1})\bigcup Q(\mathbf{i_2},\mathbf{j_2})$ on the I-line be t_1, t_2, t, respectively; and the number of vertexes on the J-line be s_1, s_2, s, respectively. Since $Q(\mathbf{i_1},\mathbf{j_1})$ and $Q(\mathbf{i_2},\mathbf{j_2})$ have common edges, we must have $t\le t_1+t_2-1$, $s\le s_1+s_2-1$.

Similar to (2.16) and (2.19), we have two bounds for $\left|\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)\right|$:

$$\left|\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)\right|\le\left(\eta T^{1/4}\right)^{8k-2(t+s-1)}\ ,\tag{2.23}$$

or

$$\left|\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)\right|\le\left(\eta T^{1/4}\right)^{8k-4t}\ .\tag{2.24}$$

For the same reason, we have also

$$\begin{aligned}\left|\mathbb{E}\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\mathbb{E}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right|&\le\left(\eta T^{1/4}\right)^{4k-2(t_1+s_1-1)+4k-2(t_2+s_2-1)}\\&<\left(\eta T^{1/4}\right)^{8k-2(t+s-1)}\ ,\end{aligned}\tag{2.25}$$

or

$$\left|\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)\right|\le\left(\eta T^{1/4}\right)^{4k-4t_1+4k-4t_2}<\left(\eta T^{1/4}\right)^{8k-4t}\ ,\tag{2.26}$$

where the last inequalities in (2.25) and (2.26) are due to the fact that $t\le t_1+t_2-1$, $s\le s_1+s_2-1$.

Since

$$\begin{aligned}
&\mathrm{Var}(m_k(A)) \\
&= \frac{1}{p^{2k+2}T^{2k}} \sum_{t,s} \sum_{M(A(t,s))} \left[\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right) - \mathbb{E}\left(\varepsilon_{Q(\mathbf{i_1},\mathbf{j_1})}\right)\mathbb{E}\left(\varepsilon_{Q(\mathbf{i_2},\mathbf{j_2})}\right)\right] \\
&:= \sum_{t,s} \widetilde{S}(t,s) \ . \qquad (2.27)
\end{aligned}$$

Using (2.23), (2.24), (2.25) and (2.26), we can bound the value of $|\widetilde{S}(t,s)|$ as follows:

$$\begin{aligned}
|\widetilde{S}(t,s)| &\le O\left(\frac{T^t p^s}{p^{2k+2}T^{2k}}\left(\eta T^{1/4}\right)^{8k-2(t+s-1)}\right) \\
&= O\left(\frac{p^{s-2k-2}}{T^{s/2-t/2-1/2}}\right) \ , \qquad (2.28)
\end{aligned}$$

or

$$\begin{aligned}
|\widetilde{S}(t,s)| &\le O\left(\frac{T^t p^s}{p^{2k+2}T^{2k}}\left(\eta T^{1/4}\right)^{8k-4t}\right) \\
&= O\left(p^{s-2k-2}\right) \ . \qquad (2.29)
\end{aligned}$$

Clearly,

$$t_1 + s_1 \le 2k+1 \ , \quad t_2 + s_2 \le 2k+1 \ ;$$

we have thus

$$t + s \le t_1 + t_2 - 1 + s_1 + s_2 - 1 \le 4k \ .$$

First, consider the case that $s > t+1$ where we use the bound in (2.28). Since

$$s - 2k - 2 - s/2 + t/2 + 1/2 = s/2 + t/2 - 2k - 3/2 \le -3/2 \ ,$$

which leads to

$$s - 2k - 2 \le -3/2 + s/2 - t/2 - 1/2 \ .$$

Combine with (2.28), we have

$$|\widetilde{S}(t,s)| \leq O\left(p^{-3/2}\left(\frac{p}{T}\right)^{\frac{1}{2}(s-t-1)}\right) \leq O\left(p^{-3/2}\right) \ . \tag{2.30}$$

Second, we use the bound in (2.29) for the case $s \leq t+1$. Recall that $t+s \leq 4k$, we have

$$s-1+s \leq t+s \leq 4k \ ,$$

which is

$$2s-1 \leq 4k \ .$$

Then, from (2.29),

$$|\widetilde{S}(t,s)| \leq O\left(p^{s-2k-2}\right) \leq O\left(p^{\frac{4k+1}{2}-2k-2}\right) = O\left(p^{-3/2}\right) \ . \tag{2.31}$$

Combine (2.27), (2.30) and (2.31), we have

$$\left|\operatorname{Var}(m_k(A))\right| \leq C(k)p^{-3/2} \ ,$$

which is summable with respect to p. Assertion (II) is then proved.

3. Convergence of the largest eigenvalue of A

In this section, we aim to show that the largest eigenvalue of A tends to 4 almost surely, which is the right edge of its LSD.

Theorem 3.1. Under the same conditions as in Theorem 2.1, with $\sup_{it}\mathbb{E}(\varepsilon_{it}^4) < \infty$ in (2.1) replaced by $\sup_{it}\mathbb{E}(|\varepsilon_{it}|^{4+\nu}) < \infty$ for some $\nu > 0$, the largest eigenvalue of A converges to 4 almost surely.

Recall that in the proof of Theorem 2.2, a main step is Lemma 2.1, which says that $|S(t,s)| \to 0$ except for one term, which is when $t=k$ and $s=k+1$. One thing to mention here is that in order to prove this lemma, k is assumed to be fixed. Then the number of

isomorphism class in $M(A(t,s))$ is a function of k, thus can be bounded by a large enough constant. So actually, we do not need to know the value of $\#\{M(A(t,s))\}$ exactly. While in the case of deriving the convergence of the largest eigenvalue, k should grow to infinity, so we can not trivially guarantee that the number of isomorphism class in $M(A(t,s))$ is still of constant order. Therefore, the main task in this section is to bound this value, making $|S(t,s)|$ ($t \neq k$ or $s \neq k+1$) still a smaller order compared with the main term $|S(k,k+1)|$ when $k \to \infty$.

Proposition 3.1. Let the conditions in Theorem 2.1 hold, with $\sup_{it} \mathbb{E}(\varepsilon_{it}^4) < \infty$ in (2.1) replaced by $\sup_{it} \mathbb{E}(|\varepsilon_{it}|^{4+\nu}) < \infty$ for some $\nu > 0$, and $k = k(p,T)$ is an integer that tends to infinity and satisfies the following conditions:

$$\begin{cases} k/\log p \to \infty, \\ kp/T \to 0, \\ k/p \to .0 \end{cases} \tag{3.1}$$

Then we have

$$\mathbb{E}(m_k(A_T)) = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot (1 + o_k(1)) \ .$$

Now suppose the above Proposition 3.1 holds true. We first show it will lead to Theorem 3.1.

Proof. (of Theorem 3.1) Using Proposition 3.1, we have the estimation that

$$\mathbb{E}(m_k(A)) = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot (1 + o_k(1)) \ , \tag{3.2}$$

then for any $\Delta > 0$, we have

$$P(l_1 > 4+\Delta) \leq P(\operatorname{tr} A^k \geq (4+\Delta)^k) \leq \frac{\mathbb{E} \operatorname{tr} A^k}{(4+\Delta)^k} = \frac{p \cdot \mathbb{E}(m_k(A))}{(4+\Delta)^k}$$

$$\leq \frac{p}{(4+\Delta)^k} \cdot \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot (1+o_k(1)) \leq \left(\frac{4p^{1/k}}{4+\Delta} \right)^k \cdot (1+o_k(1)) . \tag{3.3}$$

The right hand side tends to $\left(\frac{4}{4+\Delta}\right)^k$ since $k/\log p \to \infty$ (so $p^{1/k} \to 1$). Once we fix this $\Delta > 0$, (3.3) is summable.

The upper bound for l_1 is trivial due to our Theorem 2.2. □

Now it remains to prove our Proposition 3.1.

Proof. (of Proposition 3.1) After truncation, centralisation and rescaling, we may assume that the ε_{it}'s satisfy the condition that

$$E(\varepsilon_{it}) = 0, \ \mathrm{Var}(\varepsilon_{it}) = 1, \ |\varepsilon_{it}| \leq \delta T^{1/2} , \tag{3.4}$$

where δ is chosen such that

$$\begin{cases} \delta \to 0 \\ \delta T^{1/2-\epsilon} \to 0 \\ \delta T^{1/2} \to \infty \\ \delta^2 k \sqrt{T} \to 0 \\ \frac{kp}{\delta^2 T} \to \infty . \end{cases} \tag{3.5}$$

More detailed justifications of (3.4) are provided in Appendix B.

From the proof of Theorem 2.2, we have

$$\mathbb{E} m_k(A) = \sum_{t,s} S(t,s) = S(k,k+1) + o(1) = \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} + o(1) ,$$

where $S(k,k+1)$ is the main term that contributes to $\mathbb{E} m_k(A)$, while all other terms can be neglect. Therefore, it remains to prove that when $k \to \infty$, we still have

$$\sum_{t \neq k \text{ or } s \neq k+1} S(t,s) = \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot o_k(1) .$$

We also consider two cases:

$$\text{Case 1}: s > k+1$$

$$\text{Case 2}: s \le k+1,\ \text{but not } t = k \text{ and } s = k+1.$$

Similar to (2.16) and (2.19), we have two bounds for the expectation part:

$$\left|\mathbb{E}[\varepsilon_{j_1\,i_1}\varepsilon_{j_1\,i_2}\varepsilon_{j_2\,s+i_2}\varepsilon_{j_2\,s+i_3}\cdots\varepsilon_{j_{2k}\,s+i_{2k}}\varepsilon_{j_{2k}\,s+i_1}]\right| \le \left(\delta T^{1/2}\right)^{4k-2(t+s-1)} \tag{3.6}$$

or

$$\left|\mathbb{E}[\varepsilon_{j_1\,i_1}\varepsilon_{j_1\,i_2}\varepsilon_{j_2\,s+i_2}\varepsilon_{j_2\,s+i_3}\cdots\varepsilon_{j_{2k}\,s+i_{2k}}\varepsilon_{j_{2k}\,s+i_1}]\right| \le \left(\delta T^{1/2}\right)^{4k-4t} . \tag{3.7}$$

Consider $t=1$ first. From Wang et al. (2013), the number of isomorphism class $\#\{M(A(1,s))\}$ is bounded by

$$\begin{pmatrix} 2k \\ 2k-s \end{pmatrix} ,$$

and combine this with (2.9) and (3.6), we have

$$|S(1,s)| \le \frac{1}{p^{k+1}T^k} T p^s \left(\delta T^{1/2}\right)^{4k-2s} \begin{pmatrix} 2k \\ 2k-s \end{pmatrix} . \tag{3.8}$$

Then,

$$\begin{aligned} \Big|\sum_s S(1,s)\Big| &\le \sum_{s=1}^{2k} \frac{1}{p^{k+1}T^k} T p^s \left(\delta T^{1/2}\right)^{4k-2s} \begin{pmatrix} 2k \\ 2k-s \end{pmatrix} \\ &= \sum_{s=1}^{2k} \frac{1}{p^{k+1}T^k} T p^s \left(\delta T^{1/2}\right)^{4k-2s} \begin{pmatrix} 2k \\ s \end{pmatrix} . \end{aligned} \tag{3.9}$$

The right hand side of (3.9) can be bounded as

$$\sum_{s=1}^{2k} \frac{T}{p^{k+1}T^k} \left(\delta T^{1/2}\right)^{4k} \left(\frac{2kp}{\delta^2 T}\right)^s ,$$

which is dominated by the term when $s = 2k$ since $\frac{kp}{\delta^2 T} \to \infty$. Then (3.9) reduces to

$$\frac{1}{p^{k+1}T^k} T p^{2k} = \left(\frac{p}{T}\right)^{k-1} \to 0 . \tag{3.10}$$

Next, we consider Case 1 and Case 2 (when $t > 1$) separately. According to Wang et al. (2013), the number of isomorphism class in $M(A(s,t))$ ($t > 1$) is bounded by

$$f_{t-1}(k) \begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} , \tag{3.11}$$

where

$$f_{t-1}(k) = \frac{1}{k} \begin{pmatrix} 2k \\ t-1 \end{pmatrix} \begin{pmatrix} k \\ t \end{pmatrix} .$$

Case 1 ($s > k+1$ and $t > 1$): The part of expectation can be bounded by (3.6), and combining this with (2.9) and (3.11), we have

$$\begin{aligned} |S(t,s)| &\leq \frac{1}{p^{k+1}T^k} \sum_{M(A(t,s))} p^s T^t \cdot \left| \mathbb{E}[\varepsilon_{j_1 i_1} \varepsilon_{j_1 i_2} \cdots \varepsilon_{j_{2k}\, s+i_{2k}} \varepsilon_{j_{2k}\, s+i_1}] \right| \\ &\leq \frac{p^s T^t}{p^{k+1}T^k} \left(\delta T^{1/2}\right)^{4k-2(t+s-1)} \cdot f_{t-1}(k) \begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} . \end{aligned} \tag{3.12}$$

Since $s \geq k+2$, $t \geq 2$, and a trivial relationship that $t+s-1 \leq 2k$, we have

$$\Big| \sum_{t,s} S(t,s) \Big| \leq \sum_{t=2}^{k-1} \sum_{s=k+2}^{2k+1-t} \frac{p^s T^t}{p^{k+1}T^k} \left(\delta T^{1/2}\right)^{4k-2(t+s-1)} \cdot f_{t-1}(k) \begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} . \tag{3.13}$$

The summation over s in (3.13) can be bounded as follows:

$$\sum_{s=k+2}^{2k+1-t} \delta^{-2s} T^{-s} p^s \begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} \leq \sum_{s=k+2}^{2k+1-t} \left(\frac{2kp}{\delta^2 T} \right)^s , \tag{3.14}$$

and since $\frac{kp}{\delta^2 T} \to \infty$, the summation in (3.14) is dominated by the term of $s = 2k+1-t$. Therefore, (3.13) reduces to

$$\begin{aligned} &\sum_{t=2}^{k-1} \frac{p^{2k+1-t}T^t}{p^{k+1}T^k} \left(\delta T^{1/2}\right)^{4k-2(t+2k+1-t-1)} \cdot f_{t-1}(k) \begin{pmatrix} 2k-t \\ 2k+1-t-1 \end{pmatrix} \\ &= \sum_{t=2}^{k-1} \left(\frac{p}{T}\right)^{k-t} f_{t-1}(k) = \sum_{t=2}^{k-1} \frac{1}{k} \begin{pmatrix} 2k \\ t-1 \end{pmatrix} \begin{pmatrix} k \\ t \end{pmatrix} \left(\frac{p}{T}\right)^{k-t} . \end{aligned} \tag{3.15}$$

For the same reason, the right hand side of (3.15) inside the summation can be bounded by

$$\frac{1}{k}\left(\frac{p}{T}\right)^k\left(\frac{2k^2T}{p}\right)^t ,$$

and since $Tk^2/p = k^2/(\frac{p}{T}) \to \infty$, the dominating term in (3.15) is when $t = k-1$, which reduces to

$$\frac{1}{k}\begin{pmatrix} 2k \\ k-2 \end{pmatrix}\begin{pmatrix} k \\ k-1 \end{pmatrix}\left(\frac{p}{T}\right) = \frac{k(k-1)}{k+2}\frac{p}{T}\cdot\frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} . \tag{3.16}$$

Since $kp/T \to 0$, we have (3.16) equals

$$\frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix}\cdot o_k(1) . \tag{3.17}$$

Therefore, in this case, we have

$$\Big|\sum_{t,s} S(t,s)\Big| = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix}\cdot o_k(1) . \tag{3.18}$$

Case 2 ($2 \le t \le k$ and $s \le k+1$): For the same reason, combining the bound of the expectation part in (3.7) with (2.9) and (3.11), we have

$$\begin{aligned} |S(t,s)| &= \frac{1}{p^{k+1}T^k}\sum_{M(A(t,s))} p^sT^t\cdot\Big|\mathbb{E}[\varepsilon_{j_1\,i_1}\varepsilon_{j_1\,i_2}\cdots\varepsilon_{j_{2k}\,s+i_{2k}}\varepsilon_{j_{2k}\,s+i_1}]\Big| \\ &\le \frac{1}{p^{k+1}T^k}\left(\delta T^{1/2}\right)^{4k-4t}\cdot p^sT^t\cdot f_{t-1}(k)\begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} . \end{aligned} \tag{3.19}$$

Therefore, we have

$$\Big|\sum_{t,s} S(t,s)\Big| \le \sum_{t=2}^{k}\sum_{s=1}^{k+1} p^{s-k-1}T^{k-t}\delta^{4k-4t}f_{t-1}(k)\begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} . \tag{3.20}$$

We also consider the following three situations:

$$(1).\ t = k \text{ and } s < k+1 ,$$

$$(2).\ 1 < t < k \text{ and } s = k+1\ ,$$

$$(3).\ 1 < t < k \text{ and } s < k+1\ ,$$

and show that for all the above three situations, we have (3.20) bounded by

$$\frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot o_k(1)\ .$$

For situation (1), (3.20) reduces to

$$\sum_{s=1}^{k} p^{s-k-1} f_{k-1}(k) \begin{pmatrix} k \\ s-1 \end{pmatrix} = \sum_{s=1}^{k} p^{s-k-1} \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} \begin{pmatrix} k \\ s-1 \end{pmatrix}\ , \tag{3.21}$$

which can be bounded as

$$\sum_{s=1}^{k} p^{-k-1} \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} (kp)^s\ .$$

Therefore, the dominating term is when $s = k$, thus (3.21) reduces to

$$\frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot \frac{k}{p} = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot o_k(1)\ ,$$

which is due to the choice of k that $k/p \to 0$.

For situation (2), (3.20) reduces to

$$\begin{aligned} &\sum_{t=2}^{k-1} \delta^{4k-4t} T^{k-t} \cdot f_{t-1}(k) \begin{pmatrix} 2k-t \\ k \end{pmatrix} \\ =& \sum_{t=2}^{k-1} \delta^{4k-4t} T^{k-t} \cdot \frac{1}{k} \begin{pmatrix} 2k \\ t-1 \end{pmatrix} \begin{pmatrix} k \\ t \end{pmatrix} \begin{pmatrix} 2k-t \\ k \end{pmatrix}\ . \end{aligned} \tag{3.22}$$

Since the right hand side of (3.22) can be bounded by

$$\sum_{t=2}^{k-1} \delta^{4k} \cdot \frac{(2kT)^k}{k} \left(\frac{2k^2}{\delta^4 T}\right)^t\ , \tag{3.23}$$

which is dominated by the term of $t = k-1$ since $\frac{2k^2}{T\delta^4} = \frac{2k^2}{(\delta T^{1/4})^4} \to \infty$. Therefore, we have (3.22) bounded by

$$
\begin{aligned}
& T\delta^4 \cdot \frac{1}{k} \begin{pmatrix} 2k \\ k-2 \end{pmatrix} \begin{pmatrix} k \\ k-1 \end{pmatrix} \begin{pmatrix} k+1 \\ k \end{pmatrix} \\
= {} & \delta^4 T \frac{(k-1)k(k+1)}{k+2} \cdot \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} \\
\to {} & \frac{1}{k} \begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot o_k(1) \ ,
\end{aligned}
$$

which is due to the fact that $\delta^4 T k^2 = (\delta^2 \sqrt{T} k)^2 \to 0$.

For situation (3), we have (3.20) reduce to

$$
\begin{aligned}
& \sum_{t=2}^{k-1} \sum_{s=1}^{k} p^{s-k-1} T^{k-t} \delta^{4k-4t} \cdot f_{t-1}(k) \begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} \\
= {} & \sum_{t=2}^{k-1} \sum_{s=1}^{k} p^{s-k-1} T^{k-t} \delta^{4k-4t} \cdot \frac{1}{k} \begin{pmatrix} 2k \\ t-1 \end{pmatrix} \begin{pmatrix} k \\ t \end{pmatrix} \begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} \ . \qquad (3.24)
\end{aligned}
$$

The part of summation over s is

$$
\sum_{s=1}^{k} p^s \begin{pmatrix} 2k-t \\ s-1 \end{pmatrix} \ ,
$$

which could be bounded by

$$
\sum_{s=1}^{k} (2kp)^s \ ,
$$

therefore, the dominating term is when $s = k$. So (3.24) reduces to

$$
\sum_{t=2}^{k-1} p^{-1} \delta^{4k-4t} T^{k-t} \cdot \frac{1}{k} \begin{pmatrix} 2k \\ t-1 \end{pmatrix} \begin{pmatrix} k \\ t \end{pmatrix} \begin{pmatrix} 2k-t \\ k-1 \end{pmatrix} \ . \qquad (3.25)
$$

For the same reason, the right hand side of (3.25) can be bounded by

$$\sum_{t=2}^{k-1} p^{-1}\delta^{4k} \cdot \frac{1}{k}\left(\frac{2k^2}{T\delta^4}\right)^t (2kT)^k ,$$

which is dominated by the term of $t = k-1$ since $\frac{k^2}{T\delta^4} = \frac{k^2}{\left(\delta T^{1/4}\right)^4} \to \infty$. Therefore, (3.25) reduces to

$$\frac{T}{p}\delta^4 \cdot \frac{1}{k}\begin{pmatrix} 2k \\ k-2 \end{pmatrix}\begin{pmatrix} k \\ k-1 \end{pmatrix}\begin{pmatrix} k+1 \\ k-1 \end{pmatrix} = O\left(\frac{\delta^4 k^3 T}{p} \cdot \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix}\right) , \tag{3.26}$$

and since $\frac{\delta^4 k^3 T}{p} = (\delta^2 k\sqrt{T})^2 \cdot k/p \to 0$, we have (3.26) equals

$$\frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot o_k(1) .$$

Finally, in all the three situations, we have

$$\Big|\sum_{t,s} S(t,s)\Big| = \frac{1}{k}\begin{pmatrix} 2k \\ k-1 \end{pmatrix} \cdot o_k(1) .$$

The proof of Proposition 3.1 is complete.

□

Appendix A: Justification of truncation, centralisation and rescaling in (2.7)

A.1. Truncation

Define two $p \times T$ matrices

$$E_1 := (\varepsilon_1\ \varepsilon_2 \cdots\ \varepsilon_{T-1}\ \varepsilon_T) , \quad E_2 := (\varepsilon_{s+1}\ \varepsilon_{s+2} \cdots\ \varepsilon_{s+T-1}\ \varepsilon_{s+T}) , \tag{A.1}$$

then

$$X_T = \frac{1}{T}\sum_{t=s+1}^{s+T} \varepsilon_t \varepsilon_{t-s}^T = \frac{1}{T}E_2 E_1^T , \tag{A.2}$$

and our target matrix

$$A = \frac{T}{p} X_T X_T^T = \frac{1}{pT} E_2 E_1^T E_1 E_2^T \ . \tag{A.3}$$

Let

$$\hat{\varepsilon}_{ij} = \varepsilon_{ij} \mathbf{1}_{\{|\varepsilon_{ij}| \le \eta T^{1/4}\}} \ ,$$

$\hat{X}_T$ and $\hat{A}$ are defined by replacing all the ε_{ij} with $\hat{\varepsilon}_{ij}$ in (A.2) and (A.3).

Using Theorem A.44 in Bai and Silverstein (2010) and the inequality that

$$\text{rank}(AB - CD) \le \text{rank}(A - C) + \text{rank}(B - D) \ ,$$

we have

$$\begin{aligned}
\left\| F^A(x) - F^{\hat{A}}(x) \right\| &= \left\| F^{\frac{T}{p} X_T X_T^T}(x) - F^{\frac{T}{p} \hat{X}_T \hat{X}_T^T}(x) \right\| \\
\le \frac{1}{p} \text{rank} \left(\sqrt{\frac{T}{p}} X_T - \sqrt{\frac{T}{p}} \hat{X}_T \right) &= \frac{1}{p} \text{rank} \left(X_T - \hat{X}_T \right) \\
= \frac{1}{p} \text{rank} \left(\frac{1}{T} E_2 E_1^T - \frac{1}{T} \hat{E}_2 \hat{E}_1^T \right) &= \frac{1}{p} \text{rank} \left(E_2 E_1^T - \hat{E}_2 \hat{E}_1^T \right) \\
\le \frac{1}{p} \text{rank} \left(E_2 - \hat{E}_2 \right) + \frac{1}{p} \text{rank} \left(E_1 - \hat{E}_1 \right) & \\
= \frac{2}{p} \text{rank} \left(E_1 - \hat{E}_1 \right) \le \frac{2}{p} \sum_{i=1}^{p} \sum_{j=1}^{T} \mathbf{1}_{\{|\varepsilon_{ij}| > \eta T^{1/4}\}} \ . &
\end{aligned} \tag{A.4}$$

Since $\sup_{it} \mathbb{E}(\varepsilon_{it}^4) < \infty$, we have always

$$\frac{1}{\eta^4 pT} \sum_{i,j} \mathbb{E} \left(|\varepsilon_{ij}|^4 I_{(|\varepsilon_{ij}| > \eta T^{1/4})} \right) \longrightarrow 0 \qquad \text{as } p, T \to \infty \ .$$

Consider the expectation and variance of $\frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{T} \mathbf{1}_{\{|\varepsilon_{ij}| > \eta T^{1/4}\}}$ in (A.4):

$$\mathbb{E} \left(\frac{2}{p} \sum_{i=1}^{p} \sum_{j=1}^{T} \mathbf{1}_{\{|\varepsilon_{ij}| > \eta T^{1/4}\}} \right) \le \frac{2}{p} \sum_{i=1}^{p} \sum_{j=1}^{T} \frac{\mathbb{E} \left(|\varepsilon_{ij}|^4 \cdot \mathbf{1}_{\{|\varepsilon_{ij}| > \eta T^{1/4}\}} \right)}{\eta^4 T} = o(1) \ ,$$

$$\text{Var} \left(\frac{2}{p} \sum_{i=1}^{p} \sum_{j=1}^{T} \mathbf{1}_{\{|\varepsilon_{ij}| > \eta T^{1/4}\}} \right) \le \frac{4}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{T} \frac{\mathbb{E} \left(|\varepsilon_{ij}|^4 \cdot \mathbf{1}_{\{|\varepsilon_{ij}| > \eta T^{1/4}\}} \right)}{\eta^4 T} = o(\frac{1}{p}) \ .$$

Applying Bernstein's inequality, for all small $\varepsilon > 0$ and large p, we have

$$P\left(\frac{2}{p}\sum_{i=1}^{p}\sum_{j=1}^{T}\mathbf{1}_{\{|\varepsilon_{ij}|>\eta T^{1/4}\}} \geq \varepsilon\right) \leq 2e^{-\frac{1}{2}\varepsilon^2 p} \ . \tag{A.5}$$

Finally, combine (A.4), (A.5) with Borel-Cantelli lemma, we have with probability 1,

$$\left\| F^{A}(x) - F^{\hat{A}}(x)\right\| \to 0 \ .$$

A.2. Centralisation

Let

$$\tilde{\varepsilon}_{ij} = \hat{\varepsilon}_{ij} - \mathbb{E}\hat{\varepsilon}_{ij} \ ,$$

$\tilde{X}_T$ and $\tilde{A}$ are defined by involving the $\tilde{\varepsilon}_{ij}$'s in (A.2) and (A.3).

Similar to (A.4), we have

$$\begin{aligned}
&\left\| F^{\hat{A}}(x) - F^{\tilde{A}}(x)\right\| \leq \frac{1}{p}\text{rank}\left(\sqrt{\frac{T}{p}}\hat{X}_T - \sqrt{\frac{T}{p}}\tilde{X}_T\right) \\
=& \frac{1}{p}\text{rank}\left(\hat{E}_2\hat{E}_1^T - \tilde{E}_2\tilde{E}_1^T\right) \leq \frac{2}{p}\text{rank}\left(\hat{E}_1 - \tilde{E}_1\right) \\
=& \frac{2}{p}\text{rank}\big(\mathbb{E}(\hat{E}_1)\big) = \frac{2}{p} \to 0 \ , \quad \text{as } p \to \infty \ .
\end{aligned}$$

Therefore, we have

$$\left\| F^{\hat{A}}(x) - F^{\tilde{A}}(x)\right\| \to 0 \ .$$

A.3. Rescaling

Let

$$\sigma_{ij}^2 = \mathbb{E}\tilde{\varepsilon}_{ij}^2 \ , \quad \check{\varepsilon}_{ij} := \tilde{\varepsilon}_{ij}/\sigma_{ij} \ ,$$

then for the same reason as (A.4), we have

$$\left\| F^{\tilde{A}}(x) - F^{\check{A}}(x)\right\| \leq \frac{1}{p}\text{rank}\left(\tilde{E}_2 - \check{E}_2\right) + \frac{1}{p}\text{rank}\left(\tilde{E}_1 - \check{E}_1\right)$$

$$
\begin{aligned}
&= \frac{2}{p}\mathrm{rank}\left(\tilde{E}_1 - \check{E}_1\right) \le \frac{2}{p}\max_{ij}\left(1-\frac{1}{\sigma_{ij}}\right)\cdot \mathrm{rank}\left(\tilde{E}_1\right)\\
&\le \frac{2}{p}\max_{ij}\left(1-\frac{1}{\sigma_{ij}}\right)\min\{p,T\}\\
&= O\left(\max_{ij}\left(1-\frac{1}{\sigma_{ij}}\right)\right)\ .
\end{aligned}
$$

Since

$$
\begin{aligned}
\sigma_{ij}^2 &= \mathbb{E}\tilde{\varepsilon}_{ij}^2 = \mathbb{E}(\hat{\varepsilon}_{ij} - \mathbb{E}\hat{\varepsilon}_{ij})^2 = \mathrm{Var}(\hat{\varepsilon}_{ij}) = \mathrm{Var}(\varepsilon_{ij}\cdot \mathbf{1}_{\{|\varepsilon_{ij}|\le \eta T^{1/4}\}})\\
&\to \mathrm{Var}(\varepsilon_{ij}) = 1\ ,\ \text{as } T\to\infty\ .
\end{aligned}
$$

Therefore, we have

$$
\left\|F^{\tilde{A}}(x) - F^{\check{A}}(x)\right\| \to 0\ .
$$

Appendix B: Justification of truncation, centralisation and rescaling in (3.4)

B.1. Truncation

E_1, E_2, X_T and A are defined in (A.1), (A.2) and (A.3). Let

$$
\hat{\varepsilon}_{ij} = \varepsilon_{ij}\mathbf{1}_{\{|\varepsilon_{ij}|\le \delta T^{1/2}\}}\ ,
$$

$\hat{X}_T$ and $\hat{A}$ are defined by replacing all the ε_{ij} with $\hat{\varepsilon}_{ij}$ in (A.2) and (A.3). With the assumption that $\sup_{it}\mathbb{E}(|\varepsilon_{it}|^{4+\nu}) < \infty$, we have always

$$
\sup_{it}\frac{\mathbb{E}\left(|\varepsilon_{it}|^{4+\nu}\mathbf{1}_{\{|\varepsilon_{it}|>\delta T^{1/2}\}}\right)}{\delta^{4+\nu}} \longrightarrow 0 \quad \text{as } p,T\to\infty\ . \tag{B.1}
$$

Since

$$
A = \frac{1}{pT}E_2E_1^TE_1E_2^T\ ,
$$

whose eigenvalues are the same as those of

$$
B := \frac{1}{pT}E_1^TE_1E_2^TE_2\ ,
$$

then we have

$$
\begin{aligned}
&\left|\lambda_{\max}(A)-\lambda_{\max}(\hat{A})\right| = \left|\lambda_{\max}(B)-\lambda_{\max}(\hat{B})\right| \\
&= \left| \left\| \frac{1}{pT} E_1^T E_1 E_2^T E_2 \right\|_{op} - \left\| \frac{1}{pT} \hat{E}_1^T \hat{E}_1 \hat{E}_2^T \hat{E}_2 \right\|_{op} \right| \\
&\le \left\| \frac{1}{pT} E_1^T E_1 E_2^T E_2 - \frac{1}{pT} \hat{E}_1^T \hat{E}_1 \hat{E}_2^T \hat{E}_2 \right\|_{op} \\
&\le \left\| \frac{1}{pT} E_1^T E_1 E_2^T E_2 - \frac{1}{pT} \hat{E}_1^T \hat{E}_1 E_2^T E_2 \right\|_{op} + \left\| \frac{1}{pT} \hat{E}_1^T \hat{E}_1 E_2^T E_2 - \frac{1}{pT} \hat{E}_1^T \hat{E}_1 \hat{E}_2^T \hat{E}_2 \right\|_{op} \\
&= \left\| \frac{1}{pT} \left(E_1^T E_1 - \hat{E}_1^T \hat{E}_1 \right) E_2^T E_2 \right\|_{op} + \left\| \frac{1}{pT} \hat{E}_1^T \hat{E}_1 \left(E_2^T E_2 - \hat{E}_2^T \hat{E}_2 \right) \right\|_{op} \\
&:= J_1 + J_2 \ . \qquad\qquad \text{(B.2)}
\end{aligned}
$$

First, we have

$$
\begin{aligned}
&\left\| E_1^T E_1 - \hat{E}_1^T \hat{E}_1 \right\|_{op} = \max_{\|x\|=1} x(E_1^T E_1 - \hat{E}_1^T \hat{E}_1) x^T \\
&= \max_{\|x\|=1} \left[x(E_1^T E_1 - \hat{E}_1^T E_1) x^T + x(\hat{E}_1^T E_1 - \hat{E}_1^T \hat{E}_1) x^T \right] \\
&\le \max_{\|x\|=1} x(E_1^T E_1 - \hat{E}_1^T E_1) x^T + \max_{\|x\|=1} x(\hat{E}_1^T E_1 - \hat{E}_1^T \hat{E}_1) x^T \\
&:= J_{11} + J_{12} \ , \qquad\qquad \text{(B.3)}
\end{aligned}
$$

where

$$
\begin{aligned}
J_{11} &= \max_{\|x\|=1} x(E_1^T E_1 - \hat{E}_1^T E_1) x^T = \max_{\|x\|=1} \sum_{i,j} x_i x_j (E_1^T E_1 - \hat{E}_1^T E_1)(i,j) \\
&= \max_{\|x\|=1} \sum_{i,j} x_i x_j \sum_{k=1}^{p} \left(\varepsilon_{ki} - \hat{\varepsilon}_{ki}\right) \varepsilon_{kj} \\
&\le \max_{\|x\|=1} \sum_{k=1}^{p} \left[\Big(\sum_i x_i^2 \Big)^{1/2} \Big(\sum_i \left(\varepsilon_{ki} - \hat{\varepsilon}_{ki}\right)^2 \Big)^{1/2} \cdot \Big(\sum_j x_j^2 \Big)^{1/2} \Big(\sum_j \varepsilon_{kj}^2 \Big)^{1/2} \right] \\
&= \sum_{k=1}^{p} \left[\Big(\sum_i \left(\varepsilon_{ki} - \hat{\varepsilon}_{ki}\right)^2 \Big)^{1/2} \cdot \Big(\sum_j \varepsilon_{kj}^2 \Big)^{1/2} \right] \\
&\le \Big(\sum_{k=1}^{p} \sum_{i=1}^{T} \left(\varepsilon_{ki} - \hat{\varepsilon}_{ki}\right)^2 \Big)^{1/2} \cdot \Big(\sum_{k=1}^{p} \sum_{j=1}^{T} \varepsilon_{kj}^2 \Big)^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&= O\left(\sqrt{pT}\cdot\left(\sum_{k=1}^{p}\sum_{i=1}^{T}\varepsilon_{ki}^2\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\right)^{1/2}\right)\\
&\le O\left(\left((pT)^2\cdot\sup_{k\,i}\mathbb{E}\Big(\varepsilon_{ki}^2\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\Big)\right)^{1/2}\right)\\
&\le O\left(\left(\frac{(pT)^2}{(\delta T^{1/2})^{2+\nu}}\cdot\sup_{k\,i}\mathbb{E}\Big(|\varepsilon_{ki}|^{4+\nu}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\Big)\right)^{1/2}\right)\\
&= o\Big(\delta p T^{1/2-\nu/4}\Big)\ , \qquad \text{(B.4)}
\end{aligned}
$$

where the last inequality is due to (B.1).

For the same reason, J_{12} is also of the same order as (B.4). Therefore we have

$$
\left\|E_1^TE_1-\hat{E}_1^T\hat{E}_1\right\|_{op}\le o\Big(\delta pT^{1/2-\nu/4}\Big)\ . \qquad \text{(B.5)}
$$

Then recall the definition of J_1 in (B.2), where

$$
\begin{aligned}
J_1 &= \left\|\frac{1}{pT}\left(E_1^TE_1-\hat{E}_1^T\hat{E}_1\right)E_2^TE_2\right\|_{op}\le\frac{1}{p}\left\|E_1^TE_1-\hat{E}_1^T\hat{E}_1\right\|_{op}\cdot\frac{1}{T}\left\|E_2^TE_2\right\|_{op}\\
&\le o\Big(\delta T^{1/2-\nu/4}\Big)\to 0\ ,\quad \text{as } p,T\to\infty\ , \qquad \text{(B.6)}
\end{aligned}
$$

where the last inequality in (B.6) is due to (B.5) and the fact that $\frac{1}{T}\left\|E_2^TE_2\right\|_{op}$ is the largest eigenvalue of the sample covariance matrix $\frac{1}{T}E_2E_2^T$, which is of constant order.

For the same reason, we also have J_2 the same order as J_1, which also tends to zero. Finally, according to (B.2) we have

$$
\left|\lambda_{\max}(A)-\lambda_{\max}(\hat{A})\right|\to 0\ .
$$

B.2. Centralisation and Rescaling

Let

$$
\sigma_{it}^2=\mathrm{Var}\,\hat{\varepsilon}_{it}\ ,\quad \tilde{\varepsilon}_{it}=\frac{\hat{\varepsilon}_{it}-\mathbb{E}\hat{\varepsilon}_{it}}{\sigma_{it}}\ ,
$$

$\tilde{X}_T$ and $\tilde{A}$ are defined by replacing all the ε_{ij} with $\tilde{\varepsilon}_{ij}$ in (A.2) and (A.3). In this subsection, we will show

$$\left|\lambda_{\max}(\hat{A}) - \lambda_{\max}(\tilde{A})\right| \to 0 ,$$

which is equivalent to showing

$$\left|\lambda_{\max}(\hat{B}) - \lambda_{\max}(\tilde{B})\right| \to 0 .$$

First, since

$$\begin{aligned}
&\sup_{ki}\left|1-\sigma_{ki}^2\right| = \sup_{ki}\left|\mathbb{E}\varepsilon_{ki}^2 - \mathbb{E}\left(\varepsilon_{ki}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|\le\delta T^{1/2}\}} - \mathbb{E}\Big(\varepsilon_{ki}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|\le\delta T^{1/2}\}}\Big)\right)^2\right| \\
&= \sup_{ki}\left|\mathbb{E}\Big(\varepsilon_{ki}^2\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\Big) + \Big(\mathbb{E}\big(\varepsilon_{ki}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\big)\Big)^2\right| \\
&\le 2\cdot\sup_{ki}\left|\mathbb{E}\Big(\varepsilon_{ki}^2\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\Big)\right| \le \frac{2\cdot\sup_{ki}\mathbb{E}\Big(\left|\varepsilon_{ki}\right|^{4+\nu}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\Big)}{\big(\delta T^{1/2}\big)^{2+\nu}} \\
&= o\left(\frac{\delta^2}{T^{\frac{2+\nu}{2}}}\right) ,
\end{aligned} \tag{B.7}$$

where the last equality is due to (B.1). Finally, we have:

$$\begin{aligned}
\sup_{ki}\left|1-\frac{1}{\sigma_{ki}}\right| &= \sup_{ki}\left|\frac{\sigma_{ki}-1}{\sigma_{ki}}\right| = \sup_{ki}\left|\frac{\sigma_{ki}^2-1}{\sigma_{ki}(\sigma_{ki}+1)}\right| = O\left(\sup_{ki}\left|\sigma_{ki}^2-1\right|\right) \\
&\le o\left(\frac{\delta^2}{T^{\frac{2+\nu}{2}}}\right) ,
\end{aligned} \tag{B.8}$$

where the last inequality is due to (B.7).

Second, we have another estimation for the term $\sup_{ki}|\mathbb{E}\hat{\varepsilon}_{ki}|$ as follows:

$$\begin{aligned}
&\sup_{ki}|\mathbb{E}\hat{\varepsilon}_{ki}| = \sup_{ki}\left|\mathbb{E}\big[\varepsilon_{ki}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|\le\delta T^{1/2}\}}\big]\right| = \sup_{ki}\left|\mathbb{E}\big[\varepsilon_{ki}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\big]\right| \\
&\le \frac{\sup_{ki}\mathbb{E}\big[|\varepsilon_{ki}|^{4+\nu}\cdot\mathbf{1}_{\{|\varepsilon_{ki}|>\delta T^{1/2}\}}\big]}{\big(\delta T^{1/2}\big)^{3+\nu}} = o\left(\frac{\delta}{T^{\frac{3+\nu}{2}}}\right) .
\end{aligned} \tag{B.9}$$

Then similar to (B.2), we have

$$\left|\lambda_{\max}(\hat{B}) - \lambda_{\max}(\tilde{B})\right|$$

$$\leq \left\| \frac{1}{pT} \left(\hat{E}_1^T \hat{E}_1 - \tilde{E}_1^T \tilde{E}_1 \right) \hat{E}_2^T \hat{E}_2 \right\|_{op} + \left\| \frac{1}{pT} \tilde{E}_1^T \tilde{E}_1 \left(\hat{E}_2^T \hat{E}_2 - \tilde{E}_2^T \tilde{E}_2 \right) \right\|_{op}$$

$$:= J_3 + J_4 \ .$$

Also, similar to (B.3) and (B.4), we have

$$\left\| \hat{E}_1^T \hat{E}_1 - \tilde{E}_1^T \tilde{E}_1 \right\|_{op} = \max_{\|x\|=1} x(\hat{E}_1^T \hat{E}_1 - \tilde{E}_1^T \tilde{E}_1) x^T$$

$$\leq \max_{\|x\|=1} x(\hat{E}_1^T \hat{E}_1 - \tilde{E}_1^T \hat{E}_1) x^T + \max_{\|x\|=1} x(\tilde{E}_1^T \hat{E}_1 - \tilde{E}_1^T \tilde{E}_1) x^T$$

$$:= J_{31} + J_{32} \ , \tag{B.10}$$

with

$$J_{31} = \max_{\|x\|=1} x(\hat{E}_1^T \hat{E}_1 - \tilde{E}_1^T \hat{E}_1) x^T = \max_{\|x\|=1} \sum_{i,j} x_i x_j \sum_{k=1}^{p} \left(\hat{\varepsilon}_{ki} - \tilde{\varepsilon}_{ki} \right) \hat{\varepsilon}_{ki}$$

$$\leq \left(\sum_{k=1}^{p} \sum_{i=1}^{T} \left(\hat{\varepsilon}_{ki} - \tilde{\varepsilon}_{ki} \right)^2 \right)^{1/2} \cdot \left(\sum_{k=1}^{p} \sum_{j=1}^{T} \hat{\varepsilon}_{kj}^2 \right)^{1/2}$$

$$= O\left(\left(pT \cdot \sum_{k=1}^{p} \sum_{i=1}^{T} \left(\hat{\varepsilon}_{ki} - \tilde{\varepsilon}_{ki} \right)^2 \right)^{1/2} \right) \tag{B.11}$$

Since

$$\sum_{k=1}^{p} \sum_{i=1}^{T} \left(\hat{\varepsilon}_{ki} - \tilde{\varepsilon}_{ki} \right)^2 = \sum_{k=1}^{p} \sum_{i=1}^{T} \left(\hat{\varepsilon}_{ki} - \frac{\hat{\varepsilon}_{ki} - \mathbb{E}\hat{\varepsilon}_{ki}}{\sigma_{ki}} \right)^2$$

$$= \sum_{k=1}^{p} \sum_{i=1}^{T} \left(1 - \frac{1}{\sigma_{ki}} \right)^2 \hat{\varepsilon}_{ki}^2 + \sum_{k=1}^{p} \sum_{i=1}^{T} \frac{1}{\sigma_{ki}^2} \left(\mathbb{E}\hat{\varepsilon}_{ki} \right)^2 + \sum_{k=1}^{p} \sum_{i=1}^{T} \frac{2}{\sigma_{ki}} \left(1 - \frac{1}{\sigma_{ki}} \right) \hat{\varepsilon}_{ki} \mathbb{E}\hat{\varepsilon}_{ki}$$

$$\leq \max \left\{ O\left(pT \cdot \left(\sup_{ki} \left| 1 - \frac{1}{\sigma_{ki}} \right| \right)^2 \right), \ O\left(pT \cdot \left(\sup_{ki} \left| \mathbb{E}\hat{\varepsilon}_{ki} \right| \right)^2 \right), \right.$$

$$\left. O\left(pT \cdot \sup_{ki} \left| 1 - \frac{1}{\sigma_{ki}} \right| \cdot \sup_{ki} \left| \mathbb{E}\hat{\varepsilon}_{ki} \right| \right) \right\}$$

$$\leq \max \left\{ o\left(\frac{\delta^4 p}{T^{1+\nu}} \right), \ o\left(\frac{\delta^2 p}{T^{2+\nu}} \right), \ o\left(\frac{\delta^3 p}{T^{3/2+\nu}} \right) \right\} \ , \tag{B.12}$$

where the last inequality is due to (B.8) and (B.9). Then according to (B.11), we have the bound for the term J_{31}:

$$|J_{31}| \leq \max \left\{ o\left(\frac{\delta^2 p}{T^{\nu/2}} \right), \ o\left(\frac{\delta p}{T^{\frac{1+\nu}{2}}} \right), \ o\left(\frac{\delta \sqrt{\delta} p}{T^{1/4+\nu/2}} \right) \right\} \ . \tag{B.13}$$

For the same reason, we have the term $|J_{32}|$ can be bounded by (B.13) as well.

Therefore, we have

$$\begin{aligned}|J_3| &= \left\| \frac{1}{pT}\left(\hat{E}_1^T\hat{E}_1 - \tilde{E}_1^T\tilde{E}_1\right)\hat{E}_2^T\hat{E}_2\right\|_{op} \\ &\leq \frac{1}{p}\left\|\hat{E}_1^T\hat{E}_1 - \tilde{E}_1^T\tilde{E}_1\right\|_{op}\cdot\frac{1}{T}\left\|\hat{E}_2^T\hat{E}_2\right\|_{op} \\ &= O\left(\frac{1}{p}(J_{31}+J_{32})\right) \\ &\leq \max\left\{o\left(\frac{\delta^2}{T^{\nu/2}}\right),\ o\left(\frac{\delta}{T^{\frac{1+\nu}{2}}}\right),\ o\left(\frac{\delta\sqrt{\delta}}{T^{1/4+\nu/2}}\right)\right\} \to 0\ .\end{aligned}$$

Similar, we also have $|J_4| \to 0$, which leads to the fact that

$$\left|\lambda_{\max}(\hat{B}) - \lambda_{\max}(\tilde{B})\right| \to 0\ . \tag{B.14}$$

References

Bai, Z. D. and Yin, Y. Q. (1988a). A convergence to the semicircle law. *Ann. Probab.* **16**(2), 863-875.

Bai, Z.D. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices* (2nd edition). *Springer, 20.*

Jin, B. S., Wang, C., Bai, Z. D., Nair, K. K. and Harding, M. C. (2014). Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.* **24**(3), 1199-1225.

Lam, C. and Yao, Q.W. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.* **40**, 694–726.

Li, Z., Pan, G.M. and Yao, J. (2013). On singular value distribution of large-dimensional autocovariance matrices. *Preprint*, available at arXiv:1402.6149.

Li, Z., Wang, Q. and Yao, J. (2014). Identifying the number of factors from singular values of a large sample auto-covariance matrix *Preprint*, available at arXiv:1410.3687.

Liu, H.Y., Aue, A. and Paul, D. (2013). On the Marčenko-Pastur law for linear time series. *Preprint*, available at arXiv:1310.7270.

Tao, T. (2012). *Topics in Random Matrix Theory*. American Mathematical Society.

Wang, L. and Paul, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.* **126**, 25-52.

Wang, C., Jin, B. S., Bai, Z. D., Nair, K. K. and Harding, M. C. (2013) Strong Limit of the Extreme Eigenvalues of a Symmetrized Auto-Cross Covariance Matrix. *Preprint*, available at arXiv:1312.2277.

Wang, Q. and Yao, J. (2014) Moment approach for singular values distribution of a large auto-covariance matrix. *Preprint*, available at arXiv:1410.0752.